

ASYMPTOTIC ANALYSIS OF A DROP-PUSH MODEL FOR PERCOLATION

ELAHE ZOHOORIAN AZAD

ABSTRACT. In this article, we study a type of a one dimensional percolation model whose basic features include a sequential dropping of particles on a substrate followed by their transport via a pushing mechanism (see [S. N. Majumdar and D. S. Dean, Phys. Rev. Lett. A 11, 89 (2002)]). Consider an empty one dimensional lattice with n empty sites and periodic boundary conditions (as a necklace with n rings). Imagine then the particles which drop sequentially on this lattice, uniformly at random on one of the n sites. Letting a site can settles at most one particle, if a particle drops on an empty site, it stick there and otherwise the particle moves according to a symmetric random walk until it takes place in the first empty site it meet. We study here, the asymptotic behavior of the arrangement of empty sites and of the total displacement of all particles as well as the partial displacement of some particles.

1. INTRODUCTION

Fundamental in the domain of percolation is the manipulation of dynamic sets: sets that can grow, shrink or otherwise change over time. Some algorithms, like for example the Kruskal or Prim algorithms, for the research of the *minimal covering tree* of a graph, involve the grouping of some distinct elements into a collection of disjoint sets, and implementing two operations, UNION, that unites two sets, FIND that finds which set a given element belongs to, see [CLR90].

In a basic model, clusters with different masses change, over time, through space and when two clusters are sufficiently close they merge into a single cluster, with a probability quantified, in some sense, by a rate kernel R depending on the masses, the positions and the velocities of the two clusters. However, such a model, including the spatial distribution of clusters and their velocity, is still too complicated for analysis. A first approximation was suggested independently by Marcus [Mar68] and Lushnikov [Lus73, Lus78], considering kernels depending only on the masses of the clusters.

A Marcus–Lushnikov process [Ald99] is a continuous-time Markov process whose state space is the set of partitions of n or, equivalently, the set of measures $\mu = \sum_k \frac{n(k,t)}{n} \delta_k$ with $\sum_k kn(k,t) = n$, on the set of positive integers \mathbb{N} . The k 's stand for the sizes of clusters and $n(k,t)$ is the number of clusters with size k at time

2000 *Mathematics Subject Classification.* 68P10 (primary), 60C05, 60J65, 68R05 (secondary).

Key words and phrases. Percolation, Drop-Push model, Random Walk of Particles, Additive Coalescent, Marcus–Lushnikov Process.

t . The size- k clusters provide a fraction $\frac{k n(k,t)}{n}$ of the total size n . A Marcus-Lushnikov process evolves by instantaneous jumps according to the rule ‘each pair of clusters (c_1, c_2) merges at rate $R(c_1, c_2)/n$ ’, which R is the rate kernel of the process.

The sizes of trees in the forest of the spanning-tree model of Yao, or the sizes of blocks of cars in the classic parking model [CM04], form an additive Marcus-Lushnikov process in which the rate kernel is $R(c_1, c_2) = c_1 + c_2$. The sizes of the connected components of the random graph of Erdős-Renyi [ER60], form a multiplicative Marcus-Lushnikov process in which the rate kernel is $R(c_1, c_2) = c_1 c_2$. The average costs of the Union-Find algorithms, in the model of Erdős-Renyi, were studied by Knuth & Schönhage [KS78], and Stepanov [Ste70]. In two cases, the clusters are the connected components of a graph, and the merging of two clusters is caused by the addition of an edge between elements of these clusters. We can, in this article, suppose that the initial state consists of n cluster of size 1, which corresponds to a graph completely disconnected with n vertices but any edge. There is $n - 1$ merging between the initial state monodisperse, δ_1 , and the final state, $\frac{1}{n}\delta_n$, of the Marcus-Lushnikov process. As it soon will be seen, the model on which we work comprises the additive case: the evolution of the sizes of clusters is described here by an additive Marcus-Lushnikov process.

In first analysis, we can distinguish three different regimes in the evolution of the additive Marcus-Lushnikov processes, according to the size $A_{k,1}^n$ of the largest cluster after the k -th jump, with the interpretations concerning the fragmentation of trees [AP98, Pav77] or the analysis of hashing algorithms [CL02]: the *sparse regime* for the case in which if $\sqrt{n} = o(n - k)$, $A_{k,1}^n/n$ tends to 0 in probability; the *transition regime*, when $n - k = \Theta(\sqrt{n})$, several clusters of size $\Theta(n)$ coexist, and, once renormalized, clusters’ sizes converge to the widths of excursions of a stochastic processes related to Brownian motion; and finally the *almost full regime* for the case if $n - k = o(\sqrt{n})$, $A_{k,1}^n/n$ tends to 1 in probability, and a unique giant cluster of size $n - o(n)$ coexists with smallest clusters with total size $o(n)$.

2. MAIN THEOREMS

Considering a Marcus-Lushnikov process, at the k -th jump, two clusters with respective sizes $(S_{k,n}, s_{k,n})$, $S_{k,n} \geq s_{k,n}$ are merged, at a cost that may depend on the sizes $(S_{k,n}, s_{k,n})$. For instance, in some implementations, a label is maintained for each element, signaling the set it belongs to, and when merging two sets, one has to change the labels of the elements of one of the 2 sets. Yao [Yao76], Knuth & Schönhage [KS78], studied two algorithms *Quick-Find* and *Quick-Find-Weighted*. Quick-Find updates the labels of one of the two sets, selected arbitrarily, leading to a cumulated cost $C_{n,m}^{QF} = \sum_{k=1}^m B_{k,n}$, in which $B_{k,n} = S_{k,n}$ with probability $1/2$ and $B_{k,n} = s_{k,n}$ with probability $1/2$. Quick-Find-Weighted updates the smallest set at a cost of $c_{k,n} = s_{k,n}$, leading to a cumulated cost $C_{n,m}^{QFW} = \sum_{k=1}^m s_{k,n}$. In other contexts where coalescence of two sets occurs, costs of interest are $L_{k,n}$, the size of one of the two sets chosen randomly with a probability that is proportional to its size, i.e. $L_{k,n} = S_{k,n}$ with probability $S_{k,n}/(S_{k,n} + s_{k,n})$ and $L_{k,n} = s_{k,n}$ with probability $s_{k,n}/(S_{k,n} + s_{k,n})$.

In our model, where each particle while falling on an occupied site moves according to a symmetric random walk until it finds an empty site, the merging cost of two clusters of particles, at the dropping moment of the k -th particle, is the *movements* of this particle in the cluster on which it falls, until it finds an empty site. We indicate this movements by $M_{k,n}$, which is also the necessary time for k -th particle to find an empty site. Obviously, $M_{k,n}$ depends on the size of the corresponding cluster (see the definition of the cluster in Section 3). The partial cumulated cost is then

$$C_{n,m} = \sum_{k=1}^m M_{k,n},$$

which is interpreted as the total movements of the m first particles ($1 \leq m \leq n$). By the two following theorems, we study the concentration of the partial cost $C_{n, \lceil \alpha n \rceil}$, and the limit law of the total cost $C_{n,n-1}$, when n tends to infinity.

Theorem 2.1. *For each $\eta \in (0, 1)$, and each ε positive,*

$$\lim_n \mathbb{P} \left(\sup_{\alpha \in [0, 1-\eta]} \left| \frac{C_{n, \lceil \alpha n \rceil}}{n} - \frac{\alpha^2(\alpha^2 - 3\alpha + 3)}{6(1-\alpha)^3} \right| \geq \varepsilon \right) = 0 .$$

Theorem 2.2. *We have,*

$$\frac{C_{n,n-1}}{n^{5/2}} \xrightarrow{loi} \frac{\sqrt{2}}{6} \xi,$$

where ξ is a random variable in which the distribution is characterized by its moments:

$$\mathbb{E}(\xi^k) = \frac{k! \sqrt{\pi}}{2^{(7k-2)/2} \Gamma(\frac{5k-1}{2})} \bar{a}_k,$$

with

$$\bar{a}_k = 2(5k-6)(5k-4)\bar{a}_{k-1} + \sum_{j=1}^{k-1} \bar{a}_j \bar{a}_{k-j} \quad k \geq 2; \quad \bar{a}_1 = \sqrt{2} .$$

The suite of this article is organized in the following way: in Section 3, we explain the embedding of the additive Marcus-Lushnikov process in our model, and we calculate the probability of the merging of two arbitrary clusters. In Section 4.1, using only the properties of the symmetric random walk, we can calculate the two first moments of the partial cost, $M_{k,n}$. Theorem 2.1 is proved in Section 4.2, thanks to the convergence of the additive Marcus-Lushnikov process to the certain solutions of the Smoluchowski equation, derived by the analytical arguments in [Nor99]; we use, more precisely, Theorem 3.1 of [CM04]. In Section 5 we show that the cumulated cost of our model can be approximated by an additive functional on Cayley trees, induced by the tolls $(n^2)_{n \geq 1}$ (Proposition 1), which makes it possible to apply the results of [ZA].

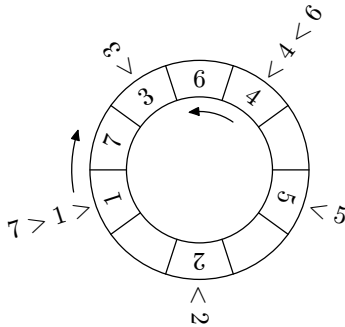


FIGURE 1. A sample of tries $f(p)$ and the resulting 3 clusters.
Here $n = 10 = 6 + 2 + 2$.

3. EMBEDDING OF THE ADDITIVE MARCUS–LUSHNIKOV PROCESS

We start with a description of the additive Marcus–Lushnikov process that helps to understand its relations to the parking scheme, generally: at the k -th step pick a first cluster P with the probability $\frac{|P|}{n}$ among the $n - k + 1$ clusters, and call it the ‘predator’ (being a size-biased pick, one obtains, generally, a cluster larger than the average cluster); then pick the ‘prey’ p uniformly among the $n - k$ remaining clusters, and suppose that P eat p , producing a unique cluster with size $|P| + |p|$. Remark that if, alternatively, both clusters are size-biased picks (resp. if both are uniform picks), we obtain the *multiplicative Marcus–Lushnikov process* (resp. the *constant kernel Marcus–Lushnikov process*, also called Kingman’s process).

Consider a lattice with n sites at a circle, on which a set $\mathcal{P} = \{1, \dots, n - 1\}$ of $n - 1$ particles drop successively and eventually stick. Letting a site can settle only one particle, each particle p drops on a random site $f(p)$. If the first chosen site $f(p)$ is on an empty site, the particle stick there. On the other hand, if the site $f(p)$ is occupied, the particle executes a symmetric simple random walk and finally it stick on the first empty site which it meets. The first chosen sites $(f(p))_{p \in \mathcal{P}}$ are assumed independent and uniform on the n sites, numbered from 1 to n .

In this model, the clusters are formed by the occupied sites, with the following conventions:

- there are as many cluster as there are empty sites,
- a cluster contains an empty site and the set of consecutive occupied sites before (going clockwise) this empty site,
- the size of the cluster is the number of sites constituting it, including the empty site,
- if an empty site follows another empty site, it is considered as a size-1 cluster of its own.

The initial configuration, with n empty sites, has thus n size-1 clusters (the *monodisperse* configuration). Each time that a particle sticks, two clusters merge, with conservation of the mass, as the empty site that disappears and the particle that replaces it both count for one mass-unit. The final configuration, once the

$n - 1$ particles are sticken, is constituted of a unique cluster with size n , and the unique empty site, uniformly distributed on $\{1, 2, \dots, n\}$. It turns out that the sizes of clusters form an additive Marcus–Lushnikov process, with kernel $K(x, y) = (x + y)/n$:

Lemma 1. *Give that k particles already sticken (that $\ell = n - k$ sites are empty), and consider then two clusters with sizes x and y . The probability that these two clusters merge at the next drop, $p_{n,k}(x, y)$, is*

$$(1) \quad \frac{x + y}{n(n - k - 1)}.$$

Proof. Let b_1, b_2, \dots, b_h ; $1 \leq h \leq k - 1$ design the non empty clusters just before the k -th particle drops (by convention, b_1 design the cluster which contains the first dropped particle). As the $k - 1$ first particles choose uniformly their sites, the order of these h clusters, from b_1 on, is a random uniform permutation. On the other side, it is not very hard to see that the empty clusters (the clusters with size 1) are merged uniformly on all their configurations:

Let l_1, l_2, \dots, l_h being the number of clusters of size 1 separating respectively the non empty clusters (i.e. the clusters of size more than 1) b_1, b_2, \dots, b_h . Thus dropping the k -th particle, two empty clusters merge conditioning that these two clusters be contiguous and the particle drops on the first cluster (clockwise), in other words, conditioning that the k -th particle drops on one of the

$$\sum_{i \in \{1, \dots, h\}} (l_i - 1)_+$$

empty sites surrounded itself by two empty sites, on the right and on the left. The conditional probability that two empty clusters merge, knowing the position of the clusters, is thus

$$\frac{1}{n} \sum_{i \in \{1, \dots, h\}} (l_i - 1)_+.$$

This probability does not depend on the sizes of the nonempty clusters, but only on their number h , and the position of the r empty clusters among the h nonempty clusters (note that $k - 1 + h + r = n$). Let us pose now $r := \sum_{i \in \{1, \dots, h\}} l_i$, the full number of empty clusters, and let us note $\mathcal{L} := \{(l_1, l_2, \dots, l_h); \sum_{i \in \{1, \dots, h\}} l_i = r\}$, the set of all configurations of (l_1, l_2, \dots, l_h) , often called *compositions* of r with h pieces. It is well-known that

$$C_h^r = \text{Card } \mathcal{L} = \binom{r + h - 1}{h - 1}.$$

Then the conditional probability that two empty clusters merge at the k -th drop, knowing the number of empty and nonempty clusters, is

$$\begin{aligned}
p_{r,h,n} &= \frac{\sum_{\mathcal{L}} \sum_{i \in \{1, \dots, h\}} (l_i - 1)_+}{nC_h^r} \\
&= \frac{1}{nC_h^r} \sum_{\mathcal{L}} [r - h + \text{Card} \{i; l_i = 0\}] \\
&= \frac{1}{nC_h^r} \left[C_h^r(r - h) + \sum_{0 \leq \ell \leq h-1} \ell C_{h-\ell}^{r-h+\ell} \binom{h}{\ell} \right] \\
&= \frac{1}{nC_h^r} [C_h^r(r - h) + C_{h-1}^r h] \\
&= \frac{2 \binom{r}{2}}{n(r + h - 1)}.
\end{aligned}$$

Remark that $\sum_{0 \leq \ell \leq h-1} \ell C_{h-\ell}^{r-h+\ell} \binom{h}{\ell}$ can be interpreted as the number of compositions of r in h pieces, such that a null piece be marked, or underlined, whereas $C_{h-1}^r h$ can be interpreted as the number of compositions of r in $h - 1$ pieces, such that one of the h interstices between the pieces be marked, or underlined: we obtain a bijective correspondence between two sets inserting one zero additional into the site of the underlined interstice, and underlining the zero so inserted. In addition, in the additive Marcus-Lushnikov model, the probability that two clusters of size 1 merge, at the stage k , if there is r pieces of size 1 and thus $h = n - k + 1 - r$ pieces of sizes higher than 1, is also

$$\binom{r}{2} \frac{2}{n(n - k)}$$

under the terms of Lemma 1.

Now let us consider the merging probability of two clusters of respective sizes $x \geq 2$ and $y \geq 2$. Let us note $N_{x,y}$ the number of empty sites met while going clockwise, from the cluster of size x to the cluster of size y : $N_{x,y}$ is uniform on $\{1, 2, \dots, n - k\}$. Obviously, if $N_{x,y} \notin \{1, n - k\}$, the two clusters are not contiguous, and cannot merge dropping of the k -th particle. For $N_{x,y} \in \{1, n - k\}$, let us note δ the exit direction of cluster, $+$ or $-$ according to whether the particle leaves there in the clockwise direction or in the opposite direction. Let us note τ the size of the cluster in which the particle drops. We have then,

$$p_{n,k}(x, y) = \frac{1}{n - k} \sum_{\delta \in \{+, -\}} \sum_{\tau \in \{x, y\}} \sum_{N_{x,y} \in \{1, n - k\}} \mathbb{P}(\delta | \tau) \frac{\tau}{n}.$$

And as (see Section 4.1)

$$\begin{aligned}
\mathbb{P}(\delta | \tau) &= \frac{\tau + 1}{2\tau} \quad \text{if } \delta = +, \\
&= \frac{\tau - 1}{2\tau} \quad \text{if } \delta = -,
\end{aligned}$$

we obtain well $p_{n,k}(x, y) = \frac{x+y}{n(n-k)}$. The merging probability of a cluster of size $x \geq 2$ with one of the r clusters of size 1 in this model, namely

$$\frac{x+1}{n} \frac{C_h^{r-1}}{C_h^r},$$

coincide also with the probability in the additive Marcus-Lushnikov model, namely $\frac{r(x+1)}{n(n-k)}$. \square

From now, assume the Marcus-Lushnikov process to be embedded in a drop particle scheme. In particular, we preserve the interpretation of $L_{j,n}$ (resp. $R_{j,n}$) as size of the cluster which is chosen by the j -th particle (resp. cluster which merges with the preceding cluster when the j -th particle sticks). We indicate by $p_{m,k}^{(j,n)}$ conditional probability that $L_{j,n}$ (which we will interpret as the size of the j -th predator before his meal) is equal to k when the cluster created by the dropping of j -th particle (the j -th predator after its meal) is of size m . According to the asymptotic behavior of $p_{m,k}^{(j,n)}$, when m is large, we hope to reach a certain intuition of the respective values of $L_{j,n}$ and $R_{j,n}$. It proves, for combinative reasons, that $p_{m,k}^{(j,n)}$ do not depend on j or n : we have, for example,

$$\begin{aligned} p_{m,k}^{(j,n)} = p_{m,k}^{(m-1,m)} &= \mathbb{P}(L_{m-1,m} = k) \\ &= \mathbb{P}(R_{m-1,m} = m - k) . \end{aligned}$$

In what follows, we shall remove thus the exponent of $p_{m,k}^{(j,n)}$.

Lemma 2.

$$p_{m,k} = \binom{m}{k} \frac{k^{k-1}(m-k)^{m-k-1}(2k-1)}{4(m-1)m^{m-1}} .$$

Proof. Let us calculate the probability $q_{m,k}^{(j,n)}$ that the j -th merge utilizes a cluster of size k and a cluster of size $m-k$, knowing that the result of the merge is of size m : as it is a probability concerning the evolution of the sizes of the clusters, it is the same one for all the models where the evolution of these sizes is described by an additive Marcus-Lushnikov process. It is thus enough to calculate $q_{m,k}^{(j,n)}$, as that was done in the parking model by Chassaing and Marchand in [CM04]. Here, we point out this calculation for the convenience of the reader. Among the n^j configurations for the j first drops, there is

$$\binom{j-1}{m-2} n m^{m-2} (n-m)^{j-m} (n-j-1)$$

configurations in which the j -th drop form a cluster of size m : there is $\binom{j-1}{m-2}$ choice for the $m-2$ other particles forming the cluster of size m , n positions for this cluster, and once the position and the particles are chosen, there is m^{m-2} ways to build this cluster of size m . The $j-m+1$ other particles can be sticken of $(n-m)^{j-m}(n-j-1)$ ways on the $n-m-1$ sites which are reserved to them.

Among the configurations in which the j -th drop form a cluster of size m , there is

$$\begin{aligned} \binom{j-1}{k-1, m-k-1} n k^{k-2} (m-k)^{m-k-2} \\ \times (k+m-k)(n-m)^{j-m} (n-j-1) \end{aligned}$$

configurations where the predator is of size k : there is $\binom{j-1}{k-1, m-k-1}$ choice for the particles of the two clusters intended to be merged, n positions for this set of two adjacent clusters, and once the particles of the two clusters and the position are chosen, the $j-1$ first particles can stick in $k^{k-2}(m-k)^{m-k-2} (n-m)^{j-m} (n-j-1)$

ways. This calculation holds if we place the cluster of size k initially, and there is then k choice for the site where the j -th particle drops. This calculation holds also if we place the cluster of size $m - k$ initially, and there is then $m - k$ choice for the site where drops j -th particle. It leads to

$$\begin{aligned} q_{m,k}^{(j,n)} &= \frac{\binom{j-1}{k-1, m-k-1} k^{k-2} (m-k)^{m-k-2} m}{\binom{j-1}{m-2} m^{m-2}} \\ &= \binom{m}{k} \frac{k^{k-1} (m-k)^{m-k-1}}{(m-1) m^{m-2}}. \end{aligned}$$

In our parking model, let x denote the probability that there exist two clusters of size k and $m - k$ side by side, at the dropping moment of the j -th particle. Then the probability that the cluster of size k is before (resp. after) the cluster of size $m - k$ is $x/2$. If the cluster of size k is on the left, the predator is of size k if the j -th particle falls on the one of its $k - 1$ occupied sites and exit from the right, with the probability

$$\frac{x(k-1)}{4n},$$

or if the j -th particle falls on the single empty site of the cluster of size k , with the probability

$$\frac{x}{2n}.$$

If the cluster of size k is on the right, the predator is of size k if the j -th particle falls on one of its $k - 1$ occupied sites and exit from the left, with the probability

$$\frac{x(k-1)}{4n}.$$

So the cluster of size k is the predator with the probability

$$\frac{xk}{2n},$$

and the cluster of size $m - k$ is the predator with the probability

$$\frac{x(m-k)}{2n}.$$

We deduced that

$$\frac{xm}{2n} = q_{m,k}^{(j,n)},$$

and that

$$p_{m,k}^{(j,n)} = \frac{k}{m} q_{m,k}^{(j,n)}.$$

Finally

$$p_{m,k} = \binom{m-1}{k-1} \frac{k^{k-1} (m-k)^{m-k-1}}{(m-1) m^{m-2}},$$

as expected. □

The Lemma 2 and the Stirling formula entails at once that

Corollary 1.

$$(2) \quad \forall k \geq 1, \lim_{m \rightarrow \infty} p_{m,m-k} = \frac{k^{k-1} e^{-k}}{k!}.$$

Thus the limiting distribution of the size of the last prey is the Borel distribution, in particular related to the explicit solutions of Smoluchowski equations [Ald99], and to the function of tree or Lambert function [Knu98]. Thus, in law, $R_{m-1,m} = \mathcal{O}(1)$. However, note that this distribution has an infinite expectation, which is coherent with the fact that $\mathbb{E}[R_{m-1,m}] = \Theta(\sqrt{m})$. What we retain of these calculations, is that provided $L_{k,n} + R_{k,n}$ be large, $R_{k,n}$ or $s_{k,n}$ be negligible compared to $L_{k,n}$.

4. PARTIAL COSTS

In this section, firstly, we calculate the first and the second moments of the movements of the k -th dropped particle, $M_{k,n}$, which we will need for the proof of Theorem 2.2. Then, we deal with the first part of the demonstration of Theorem 2.1, which is in fact a corollary of Theorem 3.1 of [CM04], stated here as Theorem 4.1.

4.1. Moments. As in the previous section, an arbitrary cluster of size s consists of $s - 1$ particles sticken successively and an empty site in the s -th position. Consider a particle drops in this cluster, on the one of these s sites chosen randomly (in the uniform way). This first choice, noted \mathcal{X}_0 , is thus a uniform random variable on $\{1, 2, \dots, s\}$. Consequently,

$$\mathbb{E}(\mathcal{X}_0) = \sum_{i=1}^s i/s = \frac{s+1}{2}$$

and

$$\mathbb{E}(\mathcal{X}_0^2) = \sum_{i=1}^s i^2/s = \frac{(s+1)(2s+1)}{6}.$$

Consider now the variable

$$\mathcal{X}_h = \mathcal{X}_0 + \sum_{i=1}^h Y_i \quad ; \quad h = 0, 1, 2, \dots,$$

representing the position of the particle in the cluster of size s , after h step. The Y_i are the Bernoulli random variables of parameter $\frac{1}{2}$ with value in $\{-1, 1\}$. We indicate by D_s the number of steps inside the cluster of size s , before the particle sticks (that one can also see as the time of receive to the edge of the cluster). We have then

$$\begin{aligned} \mathbb{P}(\mathcal{X}_{D_s} = s) &= \mathbb{P}(\mathcal{X}_{D_s} = s | \mathcal{X}_0 = s) \mathbb{P}(\mathcal{X}_0 = s) \\ &+ \mathbb{P}(\mathcal{X}_{D_s} = s | \mathcal{X}_0 \neq s) \mathbb{P}(\mathcal{X}_0 \neq s) \\ &= \frac{1}{s} + \frac{1}{2} \frac{s-1}{s} \\ &= \frac{s+1}{2s}. \end{aligned}$$

Consequently, $\mathbb{P}(\mathcal{X}_{D_s} = 0) = \frac{s-1}{2s}$. Remark that 0 indicates the last site before the cluster (in the clockwise direction), site which is empty.

The processes $\mathcal{M}_h = \mathcal{X}_h^2 - h$ and

$$\mathcal{M}'_h = \mathcal{X}_h^4 - 2(3h-2)\mathcal{X}_h^2 + h(3h-1),$$

are martingales. By stopping theorem,

$$\mathbb{E}(\mathcal{X}_{D_s}^2 - D_s) = \mathbb{E}(\mathcal{X}_0^2 - 0),$$

which gives

$$(3) \quad \mathbb{E}(D_s) = \frac{s^2 - 1}{6}.$$

Since the variables \mathcal{X}_{D_s} and D_s are not independents, the calculate of the second moment of D_s starting from \mathcal{M}'_h is not direct. To circumvent the difficulty, we add the site 0 to the cluster of size s , thus obtaining a cluster of size $s + 1$, such that we fall in a symmetrical situation: we consider then an initial position $\tilde{\mathcal{X}}_0$ uniform on $\{0, 1, \dots, s\}$. We define then \tilde{D}_s and $\tilde{\mathcal{X}}_{\tilde{D}_s}$ in a way similar to D_s and \mathcal{X}_{D_s} , but these two new variables are now independents. We have thus

$$(4) \quad \begin{aligned} \mathbb{E}(\tilde{D}_s^2) &= \mathbb{E}(\tilde{D}_s^2 | \tilde{\mathcal{X}}_0 = 0) \frac{s}{s+1} \\ &+ \mathbb{E}(\tilde{D}_s^2 | \tilde{\mathcal{X}}_0 \neq 0) \frac{s}{s+1} \\ &= \frac{s}{s+1} E(D_s^2). \end{aligned}$$

Moreover,

$$\begin{aligned} \mathbb{E}(\tilde{\mathcal{X}}_0^2) &= \frac{s(2s+1)}{6}, \\ \mathbb{E}(\tilde{\mathcal{X}}_0^4) &= \frac{6s^5 + 15s^4 + 10s^3 - s}{30(s+1)}, \\ \mathbb{P}(\tilde{\mathcal{X}}_{\tilde{D}_s} = s) &= \frac{1}{2} = \mathbb{P}(\tilde{\mathcal{X}}_{\tilde{D}_s} = 0). \end{aligned}$$

Stopping theorem for the martingale

$$\tilde{\mathcal{M}}_h = \tilde{\mathcal{X}}_h^4 - 2(3h-2)\tilde{\mathcal{X}}_h^2 + h(3h-1),$$

gives then

$$\mathbb{E}[\tilde{\mathcal{X}}_{\tilde{D}_s}^4 - 2(3\tilde{D}_s-2)\tilde{\mathcal{X}}_{\tilde{D}_s}^2 + \tilde{D}_s(3\tilde{D}_s-1)] = \mathbb{E}(\tilde{\mathcal{X}}_0^4 + \tilde{\mathcal{X}}_0^2),$$

from which it is deduced that

$$(5) \quad \mathbb{E}(D_s^2) = \frac{(s^2-1)(3s^2-7)}{45}.$$

4.2. After $\lceil \alpha n \rceil$ -th drop. Here, Theorem 4.1, gives the expression, in terms of the solution $q(k, t)$ of Smoluchowski equation, of the limit function $\varphi^\varsigma(\alpha)$ for the partial cost

$$C_{n, \lceil \alpha n \rceil}^\varsigma = \sum_{k=1}^{\lceil \alpha n \rceil} \hat{\varsigma}(S_{k,n}, s_{k,n}, U_{k,n}),$$

once $C_{n, \lceil \alpha n \rceil}^\varsigma$ is normalized by n . This theorem covers a wide class of costs, because the general expression $\hat{\varsigma}(S_{k,n}, s_{k,n}, U_{k,n})$ of the instantaneous cost of the k -th jump utilizes a randomization parameter $U_{k,n}$, uniform on $[0, 1]$. The asymptotic behavior of the partial cost is expressed according to the conditional instantaneous cost

$$\varsigma(x, y) = \mathbb{E}[\hat{\varsigma}(S_{k,n}, s_{k,n}, U_{k,n}) | (S_{k,n}, s_{k,n}) = (x, y)].$$

Theorem 4.1 requires a hypothesis little restrictive of polynomial growth of the moment of order 2 of the instantaneous conditional cost,

$$(6) \quad \forall x, y \in \mathbb{N}, \quad h(x, y) = \int_0^1 \hat{\varsigma}(x, y, u)^2 du \leq Ax^n y^m,$$

for A, m and n well selected¹. The cost $\hat{\varsigma}$ is supposed nonnegative, and $(U_{k,n})_{k \in \mathbb{N}, n \in \mathbb{N}}$ denote a sequence of independent random variables uniformly distributed on $[0, 1]$. We note φ^ς the increasing function of $[0, 1]$ in \mathbb{R}^+ defined by

$$\varphi^\varsigma(\alpha) = \int_0^{\log(\frac{1}{1-\alpha})} \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} \varsigma(k, l) \left(\frac{k+l}{2}\right) q(k, t) q(l, t) dt,$$

and

$$q(k, t) = \frac{[k(1 - e^{-t})]^{k-1} e^{-t}}{k!} \exp(-k(1 - e^{-t})).$$

We have thus

Theorem 4.1 ([CM04]). *For all $\eta > 0$,*

$$\sup_{\alpha \in [0, 1-\eta]} \left| \frac{C_{n, \lceil \alpha n \rceil}^\varsigma}{n} - \varphi^\varsigma(\alpha) \right| \xrightarrow{P} 0.$$

We pose now some notations concerning our model: $p(k)$ denotes the particle concerned with the k -th jump, the one which verifies

$$\# \{p \mid 1 \leq p \leq n-1 \text{ and } T_p \leq T_{p(k)}\} = k,$$

(where T_p indicates the moment when the particle p sticks). Let us note $f_0(p(k))$, or $f_0(k)$ to be short, the first try of $p(k)$, and note $f_j(k), j \geq 1$, the j -th try of $p(k)$ at the time of its search for an empty site. Let us note \mathcal{H}_k a σ -algebra generated by the trajectories $(f_j(\iota))_{j \geq 0, 1 \leq \iota \leq k-1}$, the particles $p(\iota); 1 \leq \iota \leq k-1$, and by $f_0(k)$. Conditioning by \mathcal{H}_k , we obtain, according to the relations (3) and (5),

$$(7) \quad \mathbb{E}[M_{k,n} | \mathcal{H}_k] = \frac{L_{k,n}^2 - 1}{6},$$

and

$$(8) \quad \mathbb{E}[M_{k,n}^2 | \mathcal{H}_k] = \frac{(L_{k,n}^2 - 1)(3L_{k,n}^2 - 7)}{45}.$$

The demonstration of Lemma 2 revealed that $L_{k,n}$ can be written

$$L_{k,n} = S_{k,n} \mathbf{1}_{V_{k,n} \leq \frac{S_{k,n}}{S_{k,n} + S_{k,n}}} + s_{k,n} \mathbf{1}_{V_{k,n} > \frac{S_{k,n}}{S_{k,n} + S_{k,n}}},$$

where $V_{k,n}$ indicates randomly a number in $[0, 1]$. To apply Theorem 4.1, we must write $M_{k,n}$ in the form

$$M_{k,n} = \hat{\varsigma}(S_{k,n}, s_{k,n}, U_{k,n}).$$

For that we must draw randomly $L_{k,n}$ in the set $\{s_{k,n}, S_{k,n}\}$, using $V_{k,n}$, as explained above, then we must randomly draw the first test of the k -th particle among the $L_{k,n}$ sites of the clusters in which it falls, for example in the form $[L_{k,n} W_{k,n}]$, where $W_{k,n}$ indicates randomly another number in $[0, 1]$, independent of $V_{k,n}$. Finally, it is necessary to simulate the random walk of the k -th particle,

¹In (6), \mathbb{N} denotes the set of strictly positive entire numbers.

for example using a sequence $(Y_{k,n,\ell})_{\ell \geq 1}$ of independent random variables $\{\pm 1\}$ symmetrical. It can be done, in a traditional way, by using the coefficients of dyadic expansion

$$U_{k,n} = \sum_{\ell \geq 1} \frac{d_{k,n,\ell}}{2^\ell}$$

to reconstitute $(V_{k,n}, W_{k,n}, (Y_{k,n,\ell})_{\ell \geq 1})$, like below

$$\begin{aligned} V_{k,n} &= \sum_{\ell \geq 1} \frac{d_{k,n,2\ell-1}}{2^\ell}, \\ W_{k,n} &= \sum_{\ell \geq 1} \frac{d_{k,n,4\ell-2}}{2^\ell}, \\ Y_{k,n,\ell} &= 2d_{k,n,4\ell} - 1. \end{aligned}$$

We have thus

$$\begin{aligned} \varsigma(x, y) &= \mathbb{E}[M_{k,n} | (S_{k,n}, s_{k,n}) = (x, y)] \\ &= \mathbb{E}[\mathbb{E}[M_{k,n} | \mathcal{H}_k] | (S_{k,n}, s_{k,n}) = (x, y)] \\ &= \frac{1}{6} \mathbb{E}[(L_{k,n}^2 - 1) | (S_{k,n}, s_{k,n}) = (x, y)] \\ &= \frac{1}{6} \left(\mathbb{E}\left[x^2 \mathbf{1}_{V_{k,n} \leq \frac{x}{x+y}} + y^2 \mathbf{1}_{V_{k,n} > \frac{x}{x+y}}\right] - 1 \right) \\ &= \frac{1}{6} \left(\frac{x^3 + y^3}{x + y} - 1 \right). \end{aligned}$$

In the same way,

$$\begin{aligned} h(x, y) &= \mathbb{E}[M_{k,n}^2 | (S_{k,n}, s_{k,n}) = (x, y)] \\ &= \mathbb{E}[(L_{k,n}^2 - 1)(3L_{k,n}^2 - 7) | (S_{k,n}, s_{k,n}) = (x, y)] \\ &= \frac{45}{(x^2 - 1)(3x^2 - 7)x + (y^2 - 1)(3y^2 - 7)y}, \\ &= \frac{45}{45(x + y)}, \end{aligned}$$

such that h be a polynomial and satisfies thus the assumption (6). We can then apply Theorem 4.1 to $\hat{\varsigma}_{k,n}$ and $\varsigma(x, y)$. Thus, for all $\eta > 0$,

$$\sup_{\alpha \in [0, 1-\eta]} \left| \frac{C_{n, [\alpha n]}}{n} - \varphi^\varsigma(\alpha) \right| \xrightarrow{P} 0,$$

with

$$\begin{aligned} \varphi^\varsigma(\alpha) &= \\ &= \frac{1}{12} \int_0^{\log(\frac{1}{1-\alpha})} \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} (k^3 + l^3 - k - l) q(k, t) q(l, t) dt \\ &= \frac{1}{6} \int_0^{\log(\frac{1}{1-\alpha})} \langle q(\cdot, t), x^3 - x \rangle \langle q(\cdot, t), 1 \rangle dt. \end{aligned}$$

To finish calculation, we use the fact that $q(\cdot, t)$ is expressed in term of the Borel distribution, or of the Lambert function T (cf. [Jan01]). More precisely, if B_a

designs a Borel random variable of parameter a , $0 < a \leq 1$, we have, for $k \geq 1$:

$$\mathbb{P}(B_a = k) = \frac{(ka)^{k-1}}{k!} e^{-ka},$$

and, for $a < 1$,

$$\begin{aligned}\mathbb{E}[B_a] &= \frac{1}{1-a}, \\ \mathbb{E}[B_a^2] &= \frac{1}{(1-a)^3}, \\ \mathbb{E}[B_a^3] &= \frac{2a+1}{(1-a)^5}.\end{aligned}$$

However it is noticed that, for $a = 1 - e^{-t}$,

$$\begin{aligned}q(k, t) &= (1-a)\mathbb{P}(B_a = k), \\ \langle q(\cdot, t), x^k \rangle &= e^{-t} \mathbb{E}[B_a^k], \\ \langle q(\cdot, t), x^3 - x \rangle &= 3e^{4t} - 2e^{3t} - 1,\end{aligned}$$

from where the calculation of $\varphi^s(\alpha)$ carried out higher, which leads to

$$\varphi^s(\alpha) = \frac{\alpha^2(\alpha^2 - 3\alpha + 3)}{6(1-\alpha)^3}.$$

5. CUMULATED COST

To obtain Theorem 2.2, it remains to estimate the error made approximating $M_{k,n}$ by $(s_{k,n} + S_{k,n})^2$ (Proposition 1), which makes it possible to reveal the relation between $C_{n,n-1}$ and the additive functional with penalties $(n^2)_{n \geq 0}$, studied in [ZA]. By keeping this goal in memory, we demonstrate initially following lemmas. We put

$$L_n = \sum_{k=1}^{n-1} L_{k,n}^2.$$

Lemma 3. $\|6C_{n,n-1} - L_n\|_2 = o(n^{5/2})$.

Proof. By developing $(6C_{n,n-1} - L_n + n - 1)^2$, we obtain:

$$\|6C_{n,n-1} - L_n - n + 1\|_2^2 = \Xi_1 + \Xi_2,$$

where

$$\Xi_1 = \sum_{k=1}^{n-1} \mathbb{E} \left[(6M_{k,n} - L_{k,n}^2 + 1)^2 \right],$$

and

$$\Xi_2 = 2 \sum_{1 \leq i < j \leq n-1} \mathbb{E} \left[(6M_{i,n} - L_{i,n}^2 + 1)(6M_{j,n} - L_{j,n}^2 + 1) \right].$$

In consequence of the relation (7), for $i < j$,

$$\mathbb{E} \left[\mathbb{E} \left[(6M_{i,n} - L_{i,n}^2 + 1)(6M_{j,n} - L_{j,n}^2 + 1) \mid \mathcal{H}_j \right] \right] = 0,$$

thus Ξ_2 disappears. According to (8), we have also

$$\mathbb{E} \left[(6M_{k,n} - L_{k,n}^2 + 1)^2 \middle| L_{k,n} \right] = \frac{7}{5} L_{k,n}^4 - 6L_{k,n}^2 + \frac{23}{5} .$$

Therefor,

$$\begin{aligned} \Xi_1 &\leq 6 \sum_{k=1}^{n-1} \mathbb{E} [L_{k,n}^4] \\ &\leq 6n^5 \int_0^1 \mathbb{E} \left[\left(\frac{L_{\lceil \alpha n \rceil, n}}{n} \right)^4 \right] d\alpha . \end{aligned}$$

According to [Pit87], for $0 < \alpha < 1$, $(B_{\lceil \alpha n \rceil, 1}^n/n)_{n \in \mathbb{N}}$ ($B_{k,1}^n$ denotes the size of the longest cluster after the k -th drop), converges in probability to 0, thus

$$\lim_n \mathbb{E} \left[\left(\frac{L_{\lceil \alpha n \rceil, n}}{n} \right)^4 \right] = 0,$$

and Lebesgue's dominated convergence Theorem completes the proof. \square

Lemma 4. $\left\| \sum_{k=1}^{n-1} (L_{k,n} + R_{k,n})^2 - L_n \right\|_1 = \mathcal{O}(n^2 \log n) .$

Proof.

$$\begin{aligned} \left\| \sum_{k=1}^{n-1} (L_{k,n} + R_{k,n})^2 - L_n \right\|_1 &= \sum_{k=1}^{n-1} \mathbb{E} [R_{k,n}^2] \\ &\quad + 2 \sum_{k=1}^{n-1} \mathbb{E} [R_{k,n} L_{k,n}] . \end{aligned}$$

Thanks to the Lemmas 4.2 and 4.8 of [CM04], for all $k \in \{1, \dots, n-1\}$ we have

$$(9) \quad \sum_{k=1}^{n-1} \mathbb{E} [R_{k,n}^2] = \mathcal{O}(n^2 \log n) ,$$

and

$$(10) \quad \mathbb{E} [R_{k,n} | L_{k,n}] = \frac{n - L_{k,n}}{n - k} .$$

Therefor,

$$\begin{aligned} n^{-2} \sum_{k=1}^{n-1} \mathbb{E} [R_{k,n} L_{k,n}] &= \int_0^1 \frac{n}{n - \lceil \alpha n \rceil} \mathbb{E} \left[\frac{L_{\lceil \alpha n \rceil, n}}{n} \right] d\alpha \\ &\quad - \int_0^1 \mathbb{E} \left[\frac{L_{\lceil \alpha n \rceil, n}^2}{n(n - \lceil \alpha n \rceil)} \right] d\alpha . \end{aligned}$$

Since for $0 < \alpha < 1$, $(B_{\lceil \alpha n \rceil, 1}^n/n)_{n \in \mathbb{N}}$ converges in probability to 0 ([Pit87]), we have

$$\lim_n \mathbb{E} \left[\frac{L_{\lceil \alpha n \rceil, n}}{n} \right] = 0,$$

and

$$\lim_n \mathbb{E} \left[\frac{L_{\lceil \alpha n \rceil, n}^2}{n(n - \lceil \alpha n \rceil)} \right] = 0,$$

and Lebesgue's dominated convergence Theorem completes the proof. \square

The two last lemmas involve the following proposition:

Proposition 1.

$$\left\| 6C_{n,n-1} - \sum_{k=1}^{n-1} (L_{k,n} + R_{k,n})^2 \right\|_1 = o\left(n^{5/2}\right).$$

Now, $\sum_{k=1}^{n-1} (L_{k,n} + R_{k,n})^2$ is precisely the additive functional on the Cayley trees induced by the penalties $(n^2)_{n \geq 0}$ studied in [ZA], to which the reader is referred for more details. Though, we represent here Theorem 1.1 of [ZA] as the following Proposition:

Proposition 2. *Let X_n be the additive functional defined on the Cayley trees, induced by the toll $(n^2)_{n \geq 0}$. Then,*

$$n^{-5/2} X_n \xrightarrow{\mathcal{L}} \sqrt{2} \xi,$$

where ξ is a random variable whose distribution is characterized by its moments.:

$$\mathbb{E}(\xi^k) = \frac{k! \sqrt{\pi}}{2^{(7k-2)/2} \Gamma(\frac{5k-1}{2})} \bar{a}_k,$$

where

$$\bar{a}_k = 2(5k-6)(5k-4)\bar{a}_{k-1} + \sum_{j=1}^{k-1} \bar{a}_j \bar{a}_{k-j}; \quad k \geq 2, \quad \bar{a}_1 = \sqrt{2}.$$

Finally Theorem 2.2 rises from Propositions 1 and 2, thanks to the following theorem [Bil95]:

Theorem 5.1. *Be $(X_n)_{n \in \mathbb{N}}$, $(Y_n)_{n \in \mathbb{N}}$ and X a random variable, such as for all n , X_n and Y_n are defined on the same probability space. If $(X_n)_{n \in \mathbb{N}}$ converges in law to X and if $(\|X_n - Y_n\|)_{n \in \mathbb{N}}$ converges in probability to 0 then $(Y_n)_{n \in \mathbb{N}}$ converges in law to X .*

ACKNOWLEDGMENTS

I wish to thank Philippe Chassaing for many fruitful discussions that helped me to solve this problem.

REFERENCES

- [Ald99] D.J. Aldous, *Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists*, Bernoulli **5** (1999), 3–48.
- [AP98] D.J. Aldous and J. Pitman, *The standard additive coalescent*, Ann. Probab. **26** (1998), 1703–1726.
- [Bil95] P. Billingsley, *Probability and measure*, John Wiley & Sons, New York, 1995.
- [CL02] P. Chassaing and G. Louchard, *Phase transition for parking blocks, Brownian excursion and coalescence*, Random Structures Algorithms **21**, no. **1** (2002), 76–119.
- [CLR90] T.H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to algorithms*, McGraw-Hill, 1990.
- [CM04] P. Chassaing and R. Marchand, *Merging costs for the additive Marcus-Lushnikov process, and union-find algorithms*, arxiv (2004), math.PR/0406094.
- [ER60] P. Erdős and A. Renyi, *On the evolution of random graphs*, Publ. Math. Inst. Hungarian Acad. Sci. **Ser. A**, **5A-2** (1960), 17–61.

- [Jan01] S. Janson, *Asymptotic distribution for the cost of linear probing hashing*, Random Structures Algorithms **19**(3-4) (2001), 438–471.
- [Knu98] D.E. Knuth, *Linear probing and graphs*, Algoritmica **22**, no.4 (1998), 561–568.
- [KS78] D.E. Knuth and A. Schönhage, *The expected linearity of a simple equivalence algorithm*, Theoret. Comput. Sci. **6**, no. 3 (1978), 281–315.
- [Lus73] A.A. Lushnikov, *Evolution of coagulating systems*, J. Colloid Interface Sci. **45** (1973), 549–556.
- [Lus78] ———, *Coagulation in finite systems*, J. Colloid Interface Sci. **65** (1978), 276–285.
- [Mar68] A.H. Marcus, *Stochastic coalescence*, Technometrics **10** (1968), 133–143.
- [Nor99] J.R. Norris, *Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent*, Ann. Appl. Probab. **9** (1999), 78–109.
- [Pav77] Yu. L. Pavlov, *The asymptotic distribution of maximum tree size in a random forest*, Th. Probab. Appl. **22** (1977), 509–5203.
- [Pit87] B. Pittel, *Linear probing: the probable largest search time grows logarithmically with the number of records*, J. Algorithms **8**, no. 2 (1987), 236–249.
- [Ste70] J. V. E. Stepanov, *The probability of the connectedness of a random graph $\mathcal{G}_m(t)$* , Teor. Verojatnost. i Primenen **15** (1970), 58–68.
- [Yao76] A. C. C. Yao, *On the average behavior of set merging algorithms*, Eighth Annual ACM Symposium on Theory of Computing (Hershey, Pa., 1976), Assoc. Comput. Mach., New York (1976), 192–195.
- [ZA] E. Zohoorian-Azad, *Limit Law of an additive functional on Cayley trees*, to appear SIAM Journal on Discrete Mathematics.

IRAN, DAMGHAN, DAMGHAN UNIVERSITY, SCHOOL OF MATHEMATICS

E-mail address: `zohorian@dubs.ac.ir`